Ergod. Th. & Dynam. Sys. (1982), **2**, 69–83
Printed in Great Britain

Geodesics on modular surfaces and continued fractions

R. MOECKEL

School of Mathematics, University of Minnesota, Minneapolis, MN 55455, *USA*

(*Received* 15 *February* 1982)

Abstract. A connection between the symbolic description of the geodesic flows on certain modular surfaces and the theory of continued fractions is developed. The well-known properties of these dynamical systems then lead to some new results about continued fractions.

Introduction

The modular group $\mathrm{Sl}(2, \mathbb{Z})$ acts on the complex plane as a group of fractional linear transformations via the correspondence

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \rightarrow \frac{az+b}{cz+d}.$$

If G is the modular group or one of its subgroups, then the action of G preserves the rational numbers and divides them into interesting equivalence classes. For example, if $G = \Gamma(2)$, the principal congruence subgroup of level 2, then there are three equivalence classes corresponding to the classification of rationals P/Q in lowest terms as odd/even, odd/odd, or even/odd.

If β is an irrational real number, then the continued fraction expansion of β leads to an infinite sequence of rational approximants P_n/Q_n which converge to β as n tends to infinity. The goal of this paper is to study the distribution of these approximants into the G-equivalence classes for typical irrationals β. The main result is proposition 2.1. One consequence of this result is that the three $\Gamma(2)$-equivalence classes occur with equal asymptotic frequency for almost every β.

The proof of proposition 2.1 depends on a connection between the theory of continued fractions and the behaviour of geodesics on the Riemann surface obtained from the upper half-plane by quotienting out the G-action. Such a connection was established in the case $G = \mathrm{Sl}(2, \mathbb{Z})$ in a classic paper of E. Artin [**2**]. In that investigation a central role was played by the tesselation of the upper half-plane induced by $\mathrm{Sl}(2, \mathbb{Z})$. A different tesselation, one more perfectly adapted to the theory of continued fractions, plays a role in our work (figure 3). The connection between this tesselation and continued fractions was known to G. Humbert as early as 1916 [**7**].

By means of this connection, number-theoretical results are found to be equivalent to results about the asymptotic behaviour of geodesics. In obtaining the latter, the ergodicity of the geodesic flow is used.

I wish to acknowledge helpful conversations with J. Moser and A. Good at ETH, Zürich. This paper was motivated by Sullivan's study of geodesic excursions on hyperbolic manifolds [9].

1. *Geodesics on modular surfaces*

Let $\mathcal{H} = \{x + iy \in \mathbb{C}: y > 0\}$ with the Poincaré metric

$$ds^2 = \frac{dx^2 + dy^2}{y^2}.$$

We refer to [1], [6] and [8] for proofs of the following basic facts about the geometry of $\mathcal{H}$.

The group

$$\mathrm{Sl}(2, \mathbb{R}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : ad - bc = 1; a, b, c, d \in \mathbb{R} \right\}$$

acts on H via the correspondence sending the matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

to the fractional linear isometry

$$z \to \frac{az + b}{cz + d}.$$

This correspondence is a group homomorphism with kernel

$$\pm \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Using this large isometry group one easily determines the geodesics in $\mathcal{H}$. The positive y-axis is a geodesic since the reflection $z \to -\bar{z}$ is an isometry and any other geodesic is the image of this one under a suitable fractional linear map, hence either another vertical line or a semicircle orthogonal to the real line. We will use the symbol $\gamma(\alpha, \beta)$, with $\alpha, \beta \in \mathbb{R} \cup \infty$, to denote the geodesic which tends to α in backward time and β in forward time (figure 1).

We are interested in the geodesic flow on $T_1\mathcal{H}$, the unit tangent bundle of $\mathcal{H}$. We describe points of $T_1\mathcal{H}$ by triples (x, y, θ) where θ denotes the angle which the unit tangent vector makes with the horizontal. The Poincaré metric induces a volume element on $T_1\mathcal{H}$ given by

$$d\mu = \frac{|dx \wedge dy \wedge d\theta|}{y^2}.$$

This is preserved both by the geodesic flow and by the action of $\mathrm{Sl}(2, \mathbb{R})$. It will be useful later to introduce coordinates (α, β, s) on $T_1\mathcal{H}$ where $\gamma(\alpha, \beta)$ is the unique geodesic tangent to the unit vector (x, y, θ) and s denotes arclength along $\gamma(\alpha, \beta)$ (figure 1). A direct computation of the Jacobian of this coordinate change shows:

$$d\mu = \frac{2|d\alpha \wedge d\beta \wedge ds|}{(\alpha - \beta)^2}. \tag{1.1}$$

https://doi.org/10.1017/S0143385700009585 Published online by Cambridge University Press

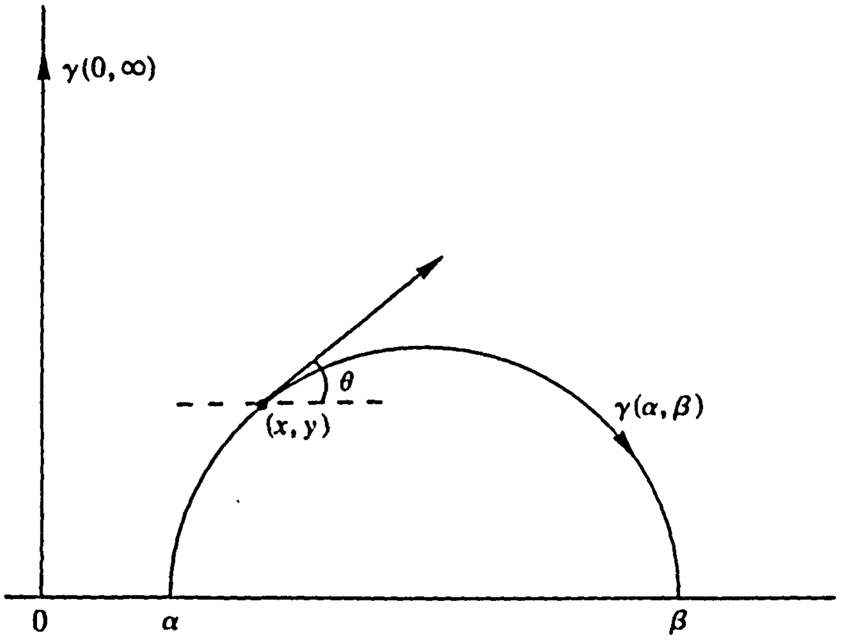

FIGURE 1.

We now consider the quotient space of $\mathcal{H}$ by the action of a Fuchsian group, i.e. a discrete subgroup of $\mathrm{Sl}(2, \mathbb{R})$. We will summarize the basic properties below [**8**]. By a fundamental region for a Fuchsian group G we mean a subset of H which aside from possible identifications of boundary points, contains exactly one representative from each G-equivalence class. Such a region which is also a geodesic polygon is called a fundamental polygon. A basic result in the theory of Fuchsian groups is that such a polygon always exists. In fact, if z_0 is any point of H not fixed by any element of G, the region

$$\mathcal{P} = \{z \in \mathcal{H}: d(z, z_0) < d(gz, z_0)\ \forall g \in G\backslash I\}$$

is the interior of a fundamental polygon. The collection of polygons $\{g(\mathcal{P}): g \in G\}$ gives a tesselation of $\mathcal{H}$. Using the fundamental polygon one can show that the quotient space $\mathcal{H}/G$ is a manifold. Since G acts as a group of isometries, there is an induced metric on $\mathcal{H}/G$. The geodesics of this metric are just the images of geodesics in $\mathcal{H}$. Now the Poincaré metric has constant curvature -1 and so also the induced metric. We also get a volume element on $T_1(\mathcal{H}/G)$. If the total volume of $T_1(\mathcal{H}/G)$ is finite then it is well known that the geodesic flow is ergodic ([**1**], [**5**], [**6**]). In fact, such flows were among the first ergodic dynamical systems known.

For applications to number theory it is natural to consider subgroups of the discrete group $\mathrm{Sl}(2, \mathbb{Z})$, the modular group. A fundamental quadrilateral $\mathcal{Q}$ for $\mathrm{Sl}(2, \mathbb{Z})$ and the associated tesselation of H are depicted in figure 2. We will use the symbol Γ for $\mathrm{Sl}(2, \mathbb{Z})$ from now on. The volume of $T_1(\mathcal{H}/\Gamma)$ is the same as that of $T_1(\mathcal{H})|_{\mathcal{Q}}$ and it is finite in spite of the fact that $\mathcal{Q}$ is not compact. If $G \subset \Gamma$ is a subgroup of finite index $[\Gamma: G] = n$, then there is a fundamental region for G made up of n copies of $\mathcal{Q}$. In fact let

$$\Gamma = GS_1 + \cdots + GS_n$$

be a coset decomposition of Γ. Then

$$S_1(\mathcal{Q}) \cup \cdots \cup S_n(\mathcal{Q})$$

https://doi.org/10.1017/S0143385700009585 Published online by Cambridge University Press

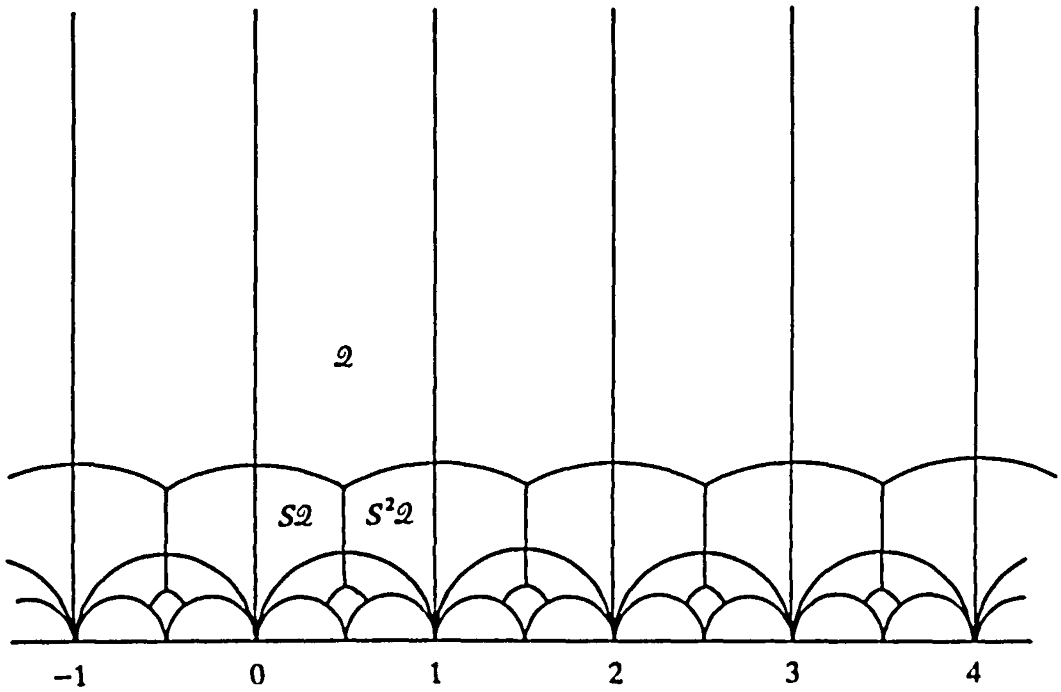

FIGURE 2

is a fundamental region for G. It follows that the volume of $T_1(\mathcal{H}/G)$ is also finite. The surfaces $\mathcal{H}/G$ with G a subgroup of finite index in Γ will be called modular surfaces.

The most familiar subgroups of Γ are the principal congruence subgroups

$$\Gamma(m)=\left\{\begin{pmatrix} a & b \\ c & d\end{pmatrix}\in\Gamma : \begin{pmatrix} a & b \\ c & d\end{pmatrix}\equiv\pm\begin{pmatrix} 1 & 0 \\ 0 & 1\end{pmatrix} \bmod m\right\}.$$

When $m=1$ the congruence condition is trivial and we find $\Gamma(1)=\Gamma$. More generally, a subgroup $G\subset\Gamma$ is called a congruence subgroup if $G\supset\Gamma(m)$ for some m. In this case there is a finite covering of $\mathcal{H}/G$ by $\mathcal{H}/\Gamma(m)$ and for our purposes such a subgroup yields nothing new. We will have occasion to consider non-congruence subgroups in § 2.

If G is a subgroup of finite index in $\Gamma=\mathrm{Sl}(2,\mathbb{Z})$, then G preserves the rational numbers $\mathbb{Q}$ splitting them into finitely many equivalence classes which we call G-cusps. It is possible to give the quotient space $(\mathcal{H}\cup\mathbb{Q})/G$ the structure of a compact surface. Therefore $\mathcal{H}/G$ is homeomorphic to a compact surface with finitely many points removed, one for each G-cusp. Cusps can be visualized as rational points in the boundary of a fundamental polygon of G. For $G=\Gamma$, every rational number is equivalent to ∞ (which we view as rational via $\infty=1/0$). Corresponding to this, ∞ is a boundary point of the fundamental quadrilateral $\mathcal{Q}$. The surface $\mathcal{H}/\Gamma$ is obtained by identifying the bounding geodesics of $\mathcal{Q}$ in the obvious way. It is a sphere with a single point removed. The Poincaré metric gives the region near ∞ a cusp-like geometry.

We will now describe a tesselation of $\mathcal{H}$ by geodesic triangles which will play a crucial role in connecting the dynamics to the number theory. By an *elementary edge* we mean a geodesic $\gamma(P/Q, P'/Q')$ whose rational endpoints satisfy $PQ'-P'Q=\pm1$. An *elementary triangle* is a triangle whose sides are elementary edges.

https://doi.org/10.1017/S0143385700009585 Published online by Cambridge University Press

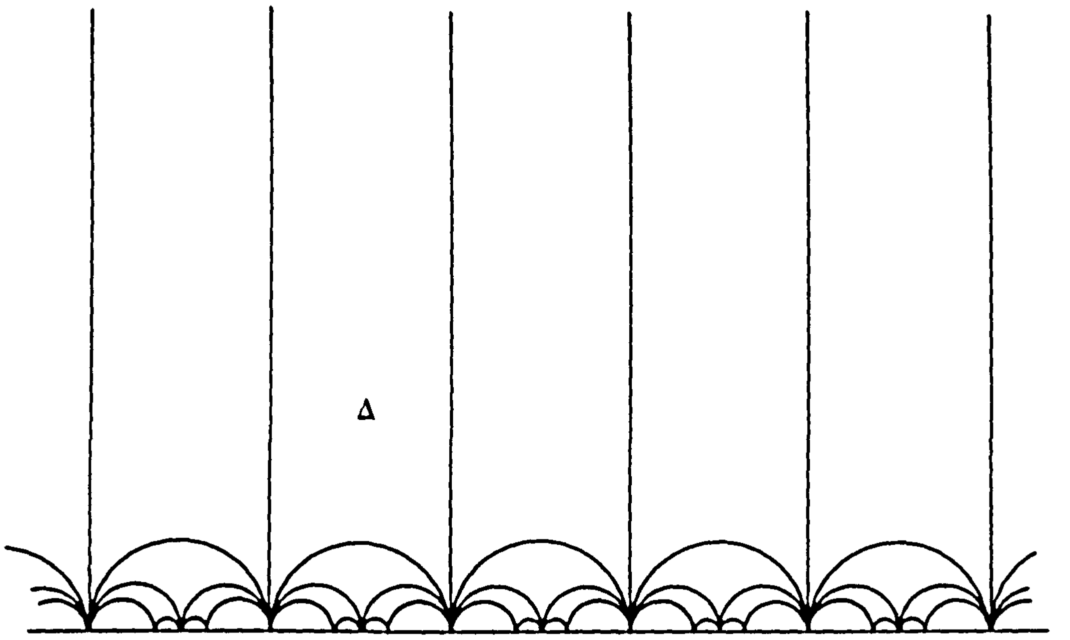

FIGURE 3

The half-plane $\mathscr{H}$ is tesselated by elementary triangles (figure 3). To see this we first note the invariance of the collection of elementary edges under the action of Γ. The image of $\gamma(P/Q, P'/Q')$ under the transformation

$$z \to \frac{az+b}{cz+d} \quad \text{is} \quad \gamma\left(\frac{aP+bQ}{cP+dQ}, \frac{aP'+bQ'}{cP'+dQ'}\right).$$

The numerators and denominators of the image rationals are the entries of

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} P & P' \\ Q & Q' \end{pmatrix}$$

which has the same determinant as

$$\begin{pmatrix} P & P' \\ Q & Q' \end{pmatrix}.$$

We also note that the triangle Δ with vertices 0, 1, and ∞ is an elementary triangle. since $\Delta \supset \mathscr{Q}$, the fundamental region for Γ, every $z \in \mathscr{H}$ lies in the image of Δ under the action of some element of Γ, so the elementary triangles cover $\mathscr{H}$. Furthermore if two distinct elementary triangles contained z, then there would be another elementary triangle which intersected $\mathscr{Q}$. But this is not the case.

There is an interesting relationship between this tesselation and the Farey sequences. The Farey sequence of order n is just the collection of all rationals P/Q with $|P| \le n$ and $|Q| \le n$. For example, the non-negative entries of the first three sequences are:

$$F_1: 0, 1, \infty$$
$$F_2: 0, \tfrac{1}{2}, 1, 2, \infty$$
$$F_3: 0, \tfrac{1}{3}, \tfrac{1}{2}, \tfrac{2}{3}, 1, \tfrac{3}{2}, 2, 3, \infty.$$

A basic property of the Farey sequences is the following: rationals $P/Q, P'/Q'$ are adjacent in the Farey sequence of order

$$\max(|P|, |Q|, |P'|, |Q'|)$$

if and only if $PQ' - P'Q = \pm 1$. For a proof see [4]. Because of this relationship we will refer to the tesselation described above as the Farey tesselation.

We will be interested in subgroups $G \subset \Gamma$ which have fundamental regions made up of elementary triangles. Let S denote the matrix

$$\begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix} \in \Gamma.$$

The corresponding fractional linear transformation has period three and maps Δ to itself. In fact,

$$\Delta = \mathcal{Q} \cup S\mathcal{Q} \cup S^2\mathcal{Q}.$$

We call a subgroup $G \subset \Gamma$ admissible if $-I \in G$, $[\Gamma : G] < \infty$, and $S \notin \Gamma^{-1}G\Gamma$.

PROPOSITION 1.1. *If G is admissible, then there is a fundamental region for G composed of $\frac{1}{3}[\Gamma : G]$ elementary triangles.*

Proof. Let $\Gamma = GS_1 + \cdots + GS_n$ be a coset decomposition of Γ where $n = [\Gamma : G]$. We let S act on the cosets from the right. The cosets GS_j, GS_jS, and GS_jS^2 are distinct for $GS_jS^a = GS_jS^b$ implies $G = GS_jS^{b-a}S_j^{-1}$ and therefore $S^{b-a} \in \Gamma^{-1}G\Gamma$. By hypothesis we must have $a = b$. Thus we have a coset decomposition

$$\Gamma = G\tilde{S}_1 + G\tilde{S}_1S + G\tilde{S}_1S^2 + \cdots + G\tilde{S}_m + G\tilde{S}_mS + G\tilde{S}_mS^2$$

with $m = \frac{1}{3}[\Gamma : G]$. As we have already remarked,

$$\tilde{S}_1\mathcal{Q} \cup \cdots \cup \tilde{S}_mS^2\mathcal{Q}$$

is a fundamental region for G. But

$$\tilde{S}_j\mathcal{Q} \cup \tilde{S}_jS\mathcal{Q} \cup \tilde{S}_jS^2\mathcal{Q} = \tilde{S}_j\Delta$$

is an elementary triangle. □

If G is admissible, the Farey tesselation induces a tesselation of $\mathcal{H}/G$ by $\frac{1}{3}[\Gamma : G]$ geodesic triangles with vertices at the cusps. If C is a cusp we write $w(C)$ for the number of triangles with vertex C and call $w(C)$ the width of the cusp. If a single triangle has two vertices at C, it contributes two to the width.

PROPOSITION 1.2. *If G is a normal subgroup then every cusp has the same width.*

Proof. The number of G-inequivalent elementary triangles with vertex ∞ is just the smallest integer k such that the matrix

$$\begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix},$$

corresponding to the translation $z \to z + k$, is an element of G. So $w(\infty) = k$. To determine $w(C)$ where C is the cusp of P/Q we map P/Q to ∞ with a fractional linear map corresponding to a matrix $T \in \Gamma$. Then $w(C)$ is the smallest k' such that

$$\begin{pmatrix} 1 & k' \\ 0 & 1 \end{pmatrix} \in TGT^{-1}.$$

Since G is normal, $TGT^{-1} = G$ and $w(C) = w(\infty)$ for every C. □

Figures 4 and 5 depict the tesselation of $\mathcal{H}/G$ for $G = \Gamma(2)$ and $G = \Gamma(3)$. In figure 4 the tesselation consists of just two triangles (front and back) and we have shown a geodesic beginning an excursion.

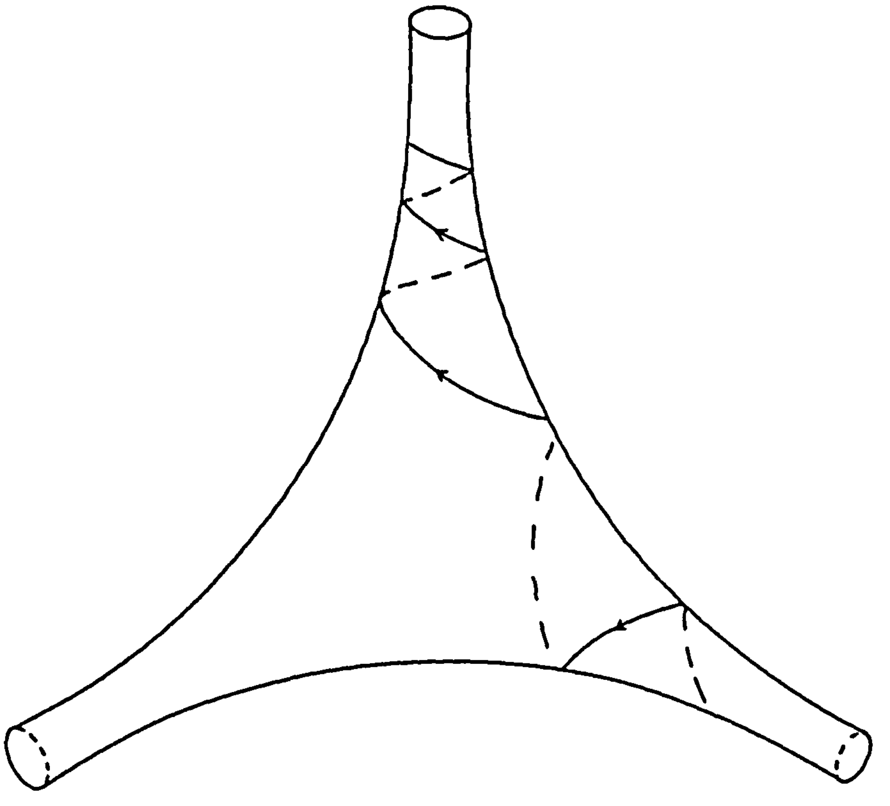

FIGURE 4

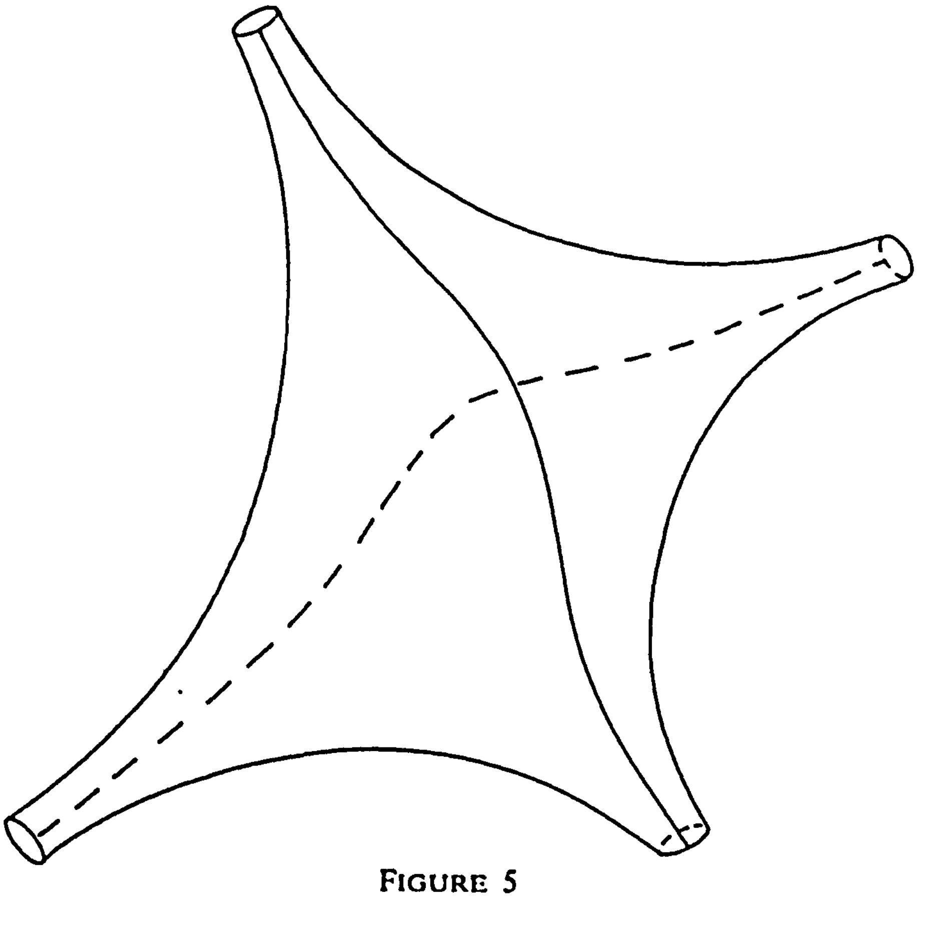

FIGURE 5

https://doi.org/10.1017/S0143385700009585 Published online by Cambridge University Press

2. *Continued fractions*

We recall some elementary facts about the continued fraction expansion of real numbers [4].

Let $b_0, b_1, b_2, \ldots$ be a sequence of integers with $b_n > 0$ for $n \geq 1$. Let $[b_0, \ldots, b_n]$ denote the expression:

$$b_0 + \cfrac{1}{b_1 + \cfrac{1}{\cdots + \cfrac{1}{b_n}}}$$

and define $[b_0, b_1, b_2, \ldots] = \lim_{n\to\infty} [b_0, \ldots, b_n]$. The limit exists and gives an irrational number β. Conversely every irrational β has a unique expansion

$$\beta = [b_0, b_1, b_2, \ldots].$$

Every rational has a similar finite expansion unique except for the fact that

$$[b_0, \ldots, b_n] = [b_0, \ldots, b_n - 1, 1]$$

for $b_n > 1$. The integer b_n is called the n'th partial quotient of β and the rational number $[b_0, \ldots, b_n]$ is called the n'th approximant to β. If we write the n'th approximant as P_n/Q_n with P_n and Q_n relatively prime integers, and if we normalize so that for positive rationals we have $P_n > 0$, $Q_n > 0$ and for negative rationals $P_n < 0$, $Q_n > 0$, then the following remarkable recursion formulae hold:

$$\left.\begin{aligned} P_{n+1} &= b_{n+1} P_n + P_{n-1}, \\ Q_{n+1} &= b_{n+1} Q_n + Q_{n-1}. \end{aligned}\right\} \tag{2.1}$$

They continue to hold if we make the conventions $P_{-1} = 1$ and $Q_{-1} = 0$ under which the -1st approximant is ∞.

We can now state our main result. Recall from § 1 that a subgroup G of $\Gamma = \mathrm{Sl}(2, \mathbb{Z})$ partitions the rational numbers into equivalence classes called G-cusps.

PROPOSITION 2.1. *Let G be an admissible subgroup of Γ and let C be a G-cusp. For almost every real number β, the rational approximants P_n/Q_n are in C with asymptotic frequency $w(C)/[\Gamma : G]$, where $w(C)$ is the width of the cusp. If G is a normal subgroup, each cusp occurs with the same asymptotic frequency for almost every β.*

By asymptotic frequency we mean

$$\lim_{N\to\infty} \frac{1}{N} |\{n \leq N : P_n/Q_n \in C\}|.$$

Before giving the proof we consider several particular subgroups G.

Under the action of

$$G = \Gamma(2) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \bmod 2 \right\}$$

two rationals in lowest terms P/Q and P'/Q' are equivalent if and only if $P \equiv P' \bmod 2$ and $Q \equiv Q' \bmod 2$. To see this we associate to a rational P/Q the integer vector

$$\begin{pmatrix} P \\ Q \end{pmatrix}.$$

Then the image of P/Q under

$$\frac{az+b}{cz+d}$$

is associated to

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} P \\ Q \end{pmatrix}$$

which is congruent to

$$\begin{pmatrix} P \\ Q \end{pmatrix} \bmod 2 \quad \text{if} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma(2).$$

On the other hand we can show that P/Q is $\Gamma(2)$-equivalent to one of $0/1$, $1/1$ or $1/0$. For example, if P/Q is of type odd/even we can find a fractional linear map

$$\frac{az+b}{cz+d}$$

taking $\infty = 1/0$ to P/Q with

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma(2).$$

This amounts to producing integers b and d such that

$$\begin{pmatrix} P & b \\ Q & d \end{pmatrix} \in \Gamma(2).$$

But since P and Q are relatively prime there exist b' and d' with $Pd' - Qb' = 1$. Reducing mod 2 we find that d is odd. The integers $b' + kP$ and $d' + kQ$ work just as well and it is easily arranged that $b = b' + kP$ be even and $d = d' + kQ$ odd. Using proposition 2.1 and the fact that $\Gamma(2)$ is normal in Γ we find that for almost every β the approximants P_n/Q_n are types odd/even, odd/odd and even/odd with asymptotic frequency $\frac{1}{3}$. As an example, we offer the golden ratio

$$1 + \sqrt{5}/2 = [1, 1, 1, \ldots]$$

whose approximants $1/0$, $1/1$, $2/1$, $3/2$, $5/3$, $8/5, \ldots$ are quotients of successive Fibonacci numbers and alternate among the three types. On the other hand,

$$1 + \sqrt{2} = [2, 2, 2, \ldots]$$

has approximants $1/0$, $2/1$, $5/2$, $12/5$, $29/12$, $70/29, \ldots$ which are never of type odd/odd.

If $G = \Gamma(m)$, $m > 2$, we find P/Q equivalent to P'/Q' if and only if

$$\begin{pmatrix} P \\ Q \end{pmatrix} \equiv \pm \begin{pmatrix} P' \\ Q' \end{pmatrix} \bmod m.$$

The numbers of cusps is therefore one-half the number of pairs (k, k') with $k, k' \in \{1, 2, \ldots, m\}$ and $gcd(k, k', m) = 1$. A rather lengthy computation shows that there are

$$\tfrac{1}{2}m^2 \prod_{\rho|m} \frac{\rho^2-1}{\rho^2}$$

cusps where the product is over primes dividing m. Since $\Gamma(m)$ is normal, each cusp occurs with frequency

$$\frac{2}{m^2} \prod_{\rho|m} \frac{\rho^2}{\rho^2-1}$$

among the rational approximants of almost every β. For example, when $m=3$ there are four cusps represented by $0/1, 1/1, 1/0, 1/2$ each occurring with frequency $1/4$. The plus or minus sign involved in determining the equivalence classes is natural since when classifying numerators and denominators mod m we should not distinguish between P/Q and $-P/-Q$.

As a final example we consider the non-congruence subgroups discovered by Fricke [3]. For each integer $m \geq 1$ we have a subgroup $G_m \subset \Gamma(2)$, the principal congruence subgroup of level 2. G_m is normal in Γ and has index $6m^2$. Furthermore $-I \in G_m$ and $S \notin G_m$ so G_m is admissible. The cusp width at ∞ (and hence at every cusp) is $2m$, so there are $3m$ cusps. To describe these we introduce continued fractions of the form:

$$a_0 - \cfrac{1}{a_1 - \cfrac{1}{a_2 \cdots \cfrac{1}{a_n}}} \tag{2.2}$$

with a_j positive or negative even integers, with the possible exception of a_n which may be odd. We will use the symbol $\langle a_0, \ldots, a_n \rangle$ for an expansion of the form (2.2). Each rational P/Q of type ∞ (odd/even) has a unique expansion $\langle a_0, \ldots, a_{2n+1} \rangle$ and a_{2n+1} is even. A rational of type 0 (even/odd) has a unique expansion $\langle a_0, \ldots, a_{2n} \rangle$ with a_{2n} even. For rationals of type 1 (odd/odd) the last partial quotient is odd and this leads to infinitely many expansions

$$\langle a_0, \ldots, a_n \rangle = \langle a_0, \ldots, a_n + 1, 2, 2, \ldots, 2, 1 \rangle.$$

We now define an index:

$$I(P/Q) = \begin{cases} a_1 + a_3 + \cdots + a_{2n+1} & \text{if type } \infty, \\ a_0 + a_2 + \cdots + a_{2n} & \text{if type } 0, \\ a_0 - a_1 + a_2 - \cdots \pm a_n & \text{if type } 1. \end{cases}$$

Note that the type 1 index does not depend on which expansion is used. The index of a type 0 or ∞ rational is even while the index of a type 1 rational is odd. The group G_M classifies the indices modulo $2m$. More precisely, the $3m$ cusps are just the sets

$$S_{\mathscr{E},k} = \{P/Q : P/Q \text{ is of type } \mathscr{E}, I(P/Q) \equiv k \bmod 2m\}$$

https://doi.org/10.1017/S0143385700009585 Published online by Cambridge University Press

where for $\mathscr{C}=0$ or ∞,

$$k\in\{2,4,\ldots,2m\}$$

and for $\mathscr{C}=1$,

$$k\in\{1,3,\ldots,2m-1\}.$$

According to proposition 2.1, each class occurs with frequency $1/3m$ among the rational approximants of almost every real number. From this it follows that the set

$$T_k=\{P/Q: I(P/Q)\equiv k \bmod 2m\}$$

occurs with frequency $\frac{1}{3}m$ if k is odd and $\frac{2}{3}m$ if k is even.

3. *Proof of proposition* 2.1

We will establish a connection between the behaviour of a geodesic $\gamma(\alpha,\beta)$ in $\mathscr{H}$ and the continued fraction expansions of α and β.

A geodesic is cut by the elementary edges into a collection of *segments*, each of which joins two edges having a common vertex. We call this common vertex the *rational number associated to the segment.* A maximal sequence of successive segments each associated to the rational P/Q will be called an *excursion into the cusp at P/Q*, and the number of segments in an excursion will be called its *size.* Figure 6 shows an excursion of size 3 into the cusp at ∞. A segment such that the preceding segment is associated to a different rational will be called an *initial segment.* Clearly every excursion begins with such a segment.

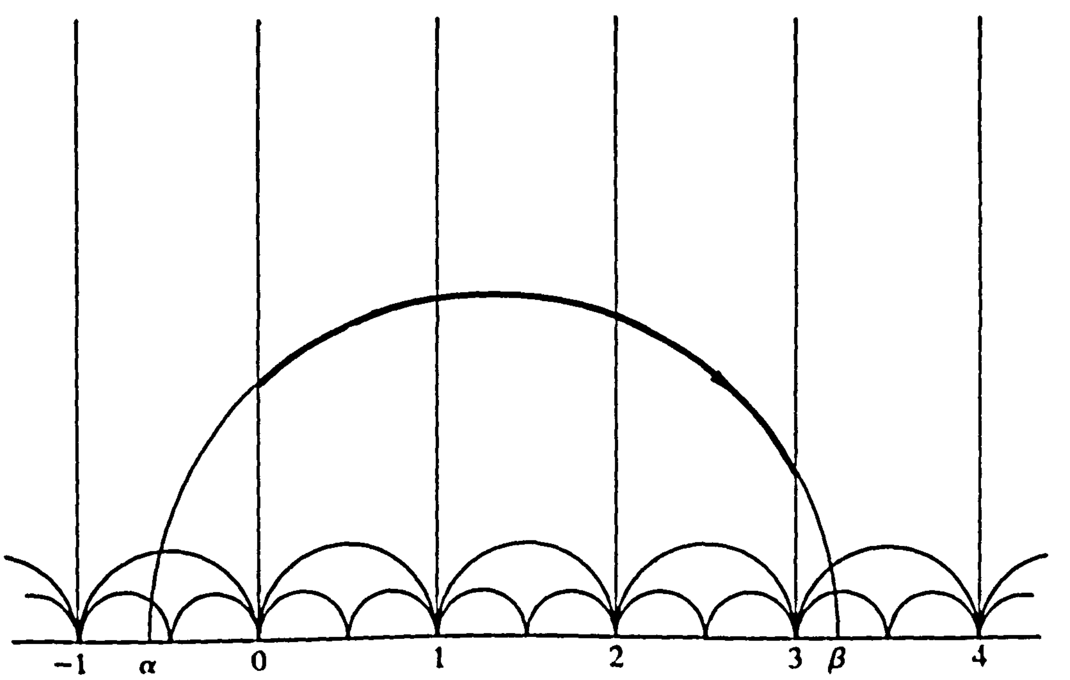

FIGURE 6

From now on we restrict attention to geodesics $\gamma(\alpha,\beta)$ with $\alpha\in[-1,0)$ and $\beta\in(1,\infty)$. Such a geodesic crosses the elementary edge $\gamma(0,\infty)$ and begins an excursion into the cusp at ∞ there. We consider this the 0'th excursion along the geodesic and we want to describe the whole sequence of excursions. To this end

https://doi.org/10.1017/S0143385700009585 Published online by Cambridge University Press

we expand β as $[b_0, b_1, b_2, \ldots]$ and set

$$[b_0, \ldots, b_n] = P_n/Q_n.$$

PROPOSITION 3.1. *The n'th excursion along $\gamma(\alpha, \beta)$ is into the cusp at P_{n-1}/Q_{n-1} and is of size b_n.*

Proof. Clearly the 0'th excursion is into the cusp at $\infty = 1/0 = P_{-1}/Q_{-1}$ and is of size $b_0 = [\beta]$ (see figure 6).

We shall need a somewhat larger class of isometries of $\mathscr{H}$ than that provided by $\mathrm{Sl}(2, \mathbb{R})$. If

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

has determinant -1 we associate to it the conjugate fractional linear map

$$z \to \frac{a\bar{z} + b}{c\bar{z} + d}$$

which is an isometry of $\mathscr{H}$. If $a, b, c, d \in \mathbb{Z}$ such a map also preserves the elementary edges and so the Farey tesselation.

We apply

$$Tz = \frac{1}{\bar{z} - b_0}.$$

T maps $\gamma(\alpha, \beta)$ to $\gamma(T\alpha, T\beta)$. But

$$T\beta = [b_1, b_2, \ldots].$$

If $-\alpha \in (0, 1]$ is expanded $[0, a_1, a_2, \ldots]$ we find

$$T\alpha = -[0, b_0, a_1, a_2, \ldots].$$

Now since T preserves the Farey tesselation, it takes excursions to excursions. The 0'th excursion maps to an excursion into $T(1/0) = T(\infty) = 0$. It follows (figure 7) that T maps the first excursion to an excursion into ∞ of size $[T\beta] = b_1$. Therefore the first excursion must have been into the cusp

$$T^{-1}\infty = b_0 = P_0/Q_0$$

as required.

For the inductive step we assume that the first n excursions were into the appropriate cusps. We apply

$$T_n z = \frac{1}{\bar{z} - b_n} \circ \frac{1}{\bar{z} - b_{n-1}} \circ \cdots \circ \frac{1}{\bar{z} - b_0},$$

which a computation using the recursion formula shows to be just

$$\frac{-Q_{n-1}\bar{z} + P_{n-1}}{Q_n\bar{z} - P_n} \quad \text{if } n \text{ is even}$$

and

$$\frac{-Q_{n-1}z + P_{n-1}}{Q_n z - P_{n-1}} \quad \text{if } n \text{ is odd.}$$

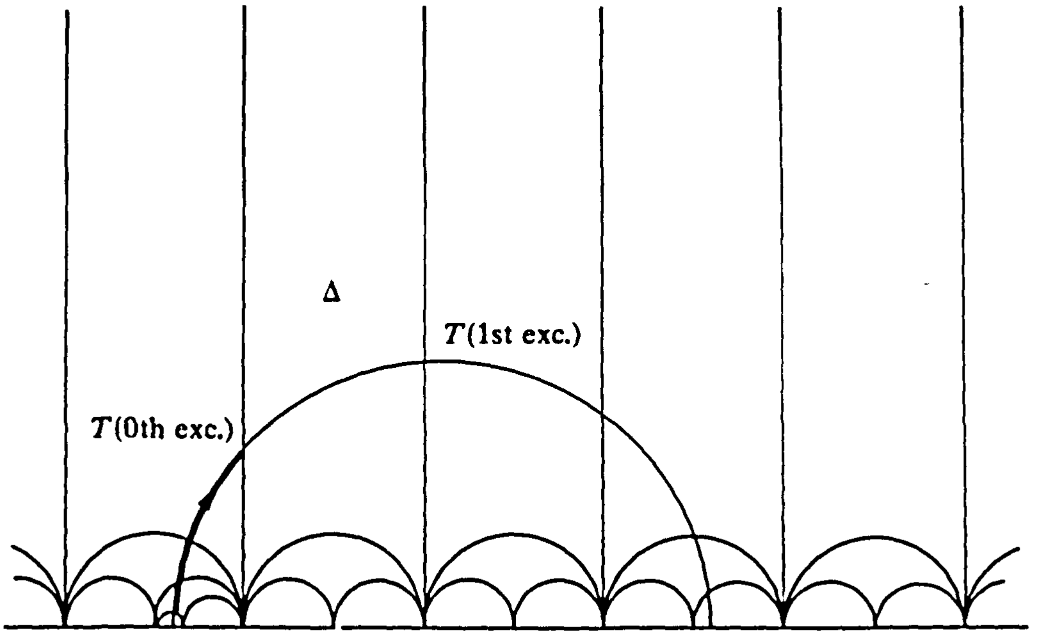

FIGURE 7

We have

$$T_n\beta = [b_{n+1}, b_{n+2}, \ldots], \qquad T_n\alpha = -[0, b_n, \ldots, b_0, a_1, a_2, \ldots],$$
$$T_n(P_{n-1}/Q_{n-1}) = 0 \quad \text{and} \quad T_n^{-1}(\infty) = P_n/Q_n.$$

Proceeding as before we conclude that the $(n+1)$'th excursion is of size b_{n+1} and is into the cusp

$$T_n^{-1}(\infty) = P_n/Q_n. \qquad \square$$

If G is an admissible subgroup of Γ, we have seen that the Farey tesselation of $\mathcal{H}$ induces a tesselation of the surface $\mathcal{H}/G$. We can define *segments* with respect to this tesselation for geodesics $\gamma(s)$ in $\mathcal{H}/G$. Each segment is associated to a G-cusp. If $\gamma(\alpha, \beta)$ is a geodesic in $\mathcal{H}$ its image in $\mathcal{H}/G$ is a geodesic, $\gamma(s)$, and a segment of $\gamma(\alpha, \beta)$ associated to P/Q maps to a segment of $\gamma(s)$ associated to the cusp containing P/Q. We define excursions and initial segments for geodesics in $\mathcal{H}/G$ to be images of excursions and initial segments in $\mathcal{H}$. It may happen that an initial segment in $\mathcal{H}/G$ is associated to the same G-cusp as the preceding segment since different rational numbers may map to the same cusp in $\mathcal{H}/G$.

It suffices to prove proposition 2.1 for almost every $\beta > 1$. To see this note that since $[\Gamma : G] < \infty$ for admissible subgroups,

$$\begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix} \in G$$

for some $k > 0$. Then the approximants P_n/Q_n of β are in a cusp C if and only if the approximants $P_n/Q_n + k$ of $\beta + k$ are in C. It is therefore sufficient to consider β in any interval of length k.

By proposition 3.1 the rational approximants P_n/Q_n to β will be in a G-cusp C with asymptotic frequency ν if and only if the image of $\gamma(\alpha, \beta)$ in $\mathcal{H}/G$ makes excursions into C with asymptotic frequency ν, for all $\alpha \in [-1, 0)$ Using formula

https://doi.org/10.1017/S0143385700009585 Published online by Cambridge University Press

(1.1) it follows easily that a set of β's of positive Lebesgue measure corresponds to a set of geodesics in $T_1(\mathcal{H}/G)$ of positive μ-measure. Consequently we need only prove:

PROPOSITION 3.2. *Almost every geodesic in $\mathcal{H}/G$ makes excursions into a cusp C with asymptotic frequency $w(C)/[\Gamma:G]$.*

Proof. Fix a triangle Δ' of the tesselation of $\mathcal{H}/G$ which has C as a vertex. We will define a function $f_{(C,\Delta')}$ on $T_1(\mathcal{H}/G)$ such that

$$\int_0^S f_{(C,\Delta')}(\gamma(s))\,ds$$

counts the number of initial segments along $\gamma(s)$ which lie in Δ', are associated to C and occur during the time interval $[0, S]$. If distinct vertices of Δ' meet at C, we define separate functions.

Now $T_1(\mathcal{H}/G)$ may be viewed as $T_1\mathcal{H}$ restricted to a fundamental region for G, and Δ' may be viewed as an elementary triangle of the fundamental region. Using the coordinates (x, y, θ) on $T_1\mathcal{H}$ we define:

$$f_{(C,\Delta')}(x, y, \theta) = \begin{cases} 1/\sigma & \text{if } (x, y, \theta) \text{ lies on an initial segment of} \\ & \text{arclength } \sigma \text{ in } \Delta', \text{ associated to } C, \\ 0 & \text{otherwise.} \end{cases}$$

Clearly the integral of $f_{(C,\Delta')}$ with respect to arclength over such an initial segment gives 1, so the function does count initial segments.

In order to use the ergodicity of the geodesic flow we must show that $f_{(C,\Delta')}$ is μ-integrable. To evaluate

$$\int_{T_1(\mathcal{H}/G)} f_{(C,\Delta')}\,d\mu$$

we note that we may assume that $\Delta' = \Delta$ and that $C = \infty$ for otherwise we can find an element of $\mathrm{Sl}(2, \mathbb{Z})$ which takes Δ' to Δ and C to ∞ while preserving initial segments, their arclengths, and the measure μ.

The segment preceding an initial segment in Δ associated to ∞ is associated to either 0 or to 1. If we integrate only over the former type we will get one-half the total integral. Initial segments of this type are of the form $\gamma(\alpha, \beta) \cap \Delta$ where $\alpha \in [-1, 0)$ and $\beta \in (1, \infty)$ (figure 6). Using formula (1.1) we find

$$\begin{aligned} \tfrac{1}{2} \int_{T_1(\mathcal{H}/G)} f_{(C,\Delta')}\,d\mu &= \int_1^\infty d\beta \int_{-1}^0 \frac{2\,d\alpha}{(\alpha-\beta)^2} \int_{\gamma(\alpha,\beta)\cap\Delta} \frac{1}{\sigma}\,ds \\ &= 2 \ln 2 < \infty \end{aligned}$$

since the innermost integral gives 1.

Now let $N_C(\gamma, S)$ be the number of excursions into C along $\gamma(s)$ begun during $[0, S]$ and let $N(\gamma, S)$ be the total number of excursions begun. Then the asymptotic frequency of excursions into C is:

$$\lim_{S\to\infty}\frac{N_C(\gamma,S)}{N(\gamma,S)}=\lim_{S\to\infty}\frac{\sum_{\Delta'}\int_0^S f_{(C,\Delta')}(\gamma(s))\,ds}{\sum_C\sum_{\Delta'}\int_0^S f_{(C,\Delta')}(\gamma(s))\,ds}$$

$$=\frac{4\ln 2w(C)}{4\ln 2\sum_C w(C)}=\frac{w(C)}{[\Gamma:G]}$$

for almost every γ. □

REFERENCES

[1] V. Arnold. *Ergodic Problems of Classical Mechanics*, Appendices 20, 21. Benjamin: Reading, Ma, 1968.

[2] E. Artin. Ein Mechanisches System mit quasiergodischen Bahnen. *Abh. Math. Sem. Hamb.* **3** (1924), 170–175.

[3] R. Fricke. Über die Substitutionsgruppen, welche zu den aus dem Legendre'schen Integralmodul $k^2(\omega)$ gezogenen Wurzeln gehören. *Math. Ann.* **28** (1887), 99–118.

[4] G. H. Hardy & E. M. Wright. *An Introduction to the Theory of Numbers.* Oxford University Press, 1960.

[5] G. Hedlund. A Metrically Transitive Group Defined by the Modular Group. *Amer. J. Math.* **57** (1935), 668–678.

[6] E. Hopf. *Ergodentheorie*, chapter V. Springer: Berlin, 1937.

[7] G. Humbert. Sur les fractions continues et les formes quadratiques binaire indéfinies. *C.R. Acad. Sci. Paris.* **162** (1916), 23–26.

[8] J. Lehner. *A Short Course in Automorphic Functions.* Holt, Rinehart, and Winston, New York, 1966.

[9] D. Sullivan. Disjoint spheres, approximation by imaginary quadratic numbers, and the logarithm law for geodesics. (Preprint.)

https://doi.org/10.1017/S0143385700009585 Published online by Cambridge University Press